

FFCUL/LASIGE TEAM FOURTH YEAR INTERIM REPORT

The Epiwork project started in February 2009 and will run for 48 months (a 6 months extension has been requested). According to the work plan FFCUL is involved in the following work packages:

- WP 1 Population Models and Contact Networks
- WP 3 Information platform
- WP 4 Epidemic Modeling Platform
- WP 7 Management

The FFCUL participation in Epiwork involves two groups:

- CMAF Group
- LASIGE Group.

This report summarises the progress in the fifth semester of the project activity by the LASIGE Group, which leads WP3 and participates in WP4 and WP7.

In the 7th reporting period (1st semester of the 4th year), the commitment was 35.49 persons.month, 31.18 pm from technicians, and 4.30 pm from permanent staff.

The LASIGE Team working on WP3 and W4 in this period includes:

- Mário J. Silva (IST Faculty, worked in the reporting period in WP3, WP4 and WP7)
- Francisco Couto (FCUL Faculty, worked in the reporting period in WP3)
- Dulce Domingos (FCUL Faculty, worked in the reporting period in WP3 and WP4)
- Paulo Graça (software developer, working full-time in WP3 and WP4)
- João Zamite (Graduate student, worked full-time in the reporting period).
- João Ferreira (Graduate student, worked full-time in the reporting period).
- Carlos Sousa (Graduate student, worked full-time in the reporting period)
- Tiago Posse (Graduate student, worked full-time in the reporting period)
- Vera Carvalho (Undergraduate student, started working full-time on Jul 13)
- Hugo Sousa (Undergraduate student, started working full-time on Jul 13)

The following changes occurred in the FFCUL/LASIGE Team:

There is a discrepancy in the profiles of the people hired to work on Epiwork at LASIGE and the initially planned, the reason being that we were unable to fill the software engineer and a post-doc positions as planned, with previous directly related experience and salaries matching our budget.

As a result, we started using Informatics Engineering and Biomedical Informatics students on the infrastructure setup assisted by FCUL faculty and senior technical staff. LASIGE has hired more people in the first two years at a reduced cost, which was the contingency measure found to manage the situation.

Two new undergraduate students joined Epiwork at the end of the reporting period to work on WP3, Vera Carvalho and Hugo Santos. Vera Carvalho is surveying epidemiology research papers to identify epidemiologic resources to be cataloged in the Epidemic Marketplace. Hugo Santos on the other hand is creating an epidemiology semantic network to be revised by Miquel Portas, the author of the Dictionary of Epidemiology, which we will use in the Epidemic Marketplace.

The FFCUL/LASIGE Team has been working on adapting its planning and to incorporate a six-months extension of the project. This will guarantee support for another 6 months of the Epidemic Marketplace along with its continued improvement with better performance, reliability and additional functionalities. This will entail not only an increase of the total manpower to commit to the project but also an extra allocation of funds for travel and consumables. The funds for the additional expenses will come from a reshuffling of the budget, where we have been under-spending as a result in the discrepancy of the hired profiles above reported.

1. WP3 — Information platform: Work progress, collaborations and achievements during the period

This Work Package is lead by FFCUL, with a total contribution of 82 persons.month (60 hired technicians + (22) academics).

2.1 Background Information on WP3

Work package number	3	Start date or starting event:				Month 1		
Work package title	Information platform							
Activity type	RTD							
Participant number	12	1	2	3	4	5	9	10
Participant short name	FFCUL	ISI	FGC-IGC	TAU	MPI-DS	AIBV	BIU	FBK-IRST
Person-months per participant	82	60	4	8	42	2	11	6

The whole WP3 activity is structured into four tasks:

Task 3.1 – Data Collection.

Participants: FFCUL, ISI, FBK-IRST, BIU, MPI-DS, FGC-IGC, AIBV.

Description: Realistic simulations of epidemic processes crucially depend on the availability of datasets describing human behaviour and pathogen-host interactions. Datasets include population movement data, social and behavioural data, health related data, geographic data, detailed geo-temporal epidemic incidence and immunization data, pathogen evolution and multi-strains circulation data. Data can come from a variety of different

sources, including hospital records, country statistics, Web content, and others. It can range from a global scale, such as the worldwide air transportation infrastructure, down to the detailed description of individual **activities** at a minute-by-minute scale. This task will create a catalogue of databases of epidemiological data across Europe, with extensive meta-data describing the main characteristics of the available information sources. This catalogue will be integrated with a collaborative platform that will be set up for online discussion and exchange of meta-data among the participants.

Task 3.2 – Meta-Model Design.

Participants: FFCUL, IGC, ISI.

Description: While some of the previously mentioned datasets are freely available on the Web (e.g. WHO Global Health Atlas, Eurostat), they are often scattered in different repositories, cover partial regions of the world and come in different formats, according to different standards and classifications. The project envisions a unified and integrated approach for the management of these resources, with the design and implementation of an Epidemic Marketplace Platform, publicly available on the web. The platform supports the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. It is an on-line social networking site that will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on a Web2.0 approach, users will become active participants, sharing information and data, and collaborating online, rather than being satisfied with a passive information consumer/viewer role. We envision proposing a simple reference format, which will facilitate the navigation and use of the datasets. Each dataset will come with a metadata file, signaling general metadata for resource management, containing data such as: the title, the date of submission, version, the source of the data and coverage. Moreover, the metadata will include information for a more thorough description of the data included in the dataset, providing a framework for a more specific description, for example, of epidemiologic and geographic data. The Marketplace will support flexible and intuitive tools for navigation and selection of resources. Standard classifications as well as tagging systems proposed by users will be supported.

Task 3.3 – Epidemic Marketplace Platform.

Participants: FFCUL, ISI.

Description: This task will implement a platform based on the integration of grid technology and publicly available services and software on the web to support the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. The Epidemic Marketplace Platform will be an on-line social networking site that

will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on some of the Web2.0 characteristics, users will become active participants, generating information and providing data for sharing, and collaborating online, rather than being satisfied with a passive information consumer/viewer role. More specifically, researchers can use and contribute to the Marketplace in several different ways. They can: (1) use it as a catalogue of data sources containing the metadata describing existing databases; (2) view, download, tag, and comment on the available resources; (3) provide compliant datasets and relevant information; (4) use it as a forum where to publish information about their own data, seek modellers to collaborate with, share and distribute their new findings.

Task 3.4 – Evaluation and monitoring of the use of the catalogue and collaboration services.

Participant: FFCUL.

Description: This task involves the monitoring of epidemiological data exchanges performed through the mediating services platform. The evaluation will assess not only the coverage of the catalogued resources, but the users' satisfaction with the user interface and integrated collaborative tools made available through the epidemiological marketplace platform. More importantly, the analysis of the collected datasets and their annotations and usage will provide a rich environment for deriving an epidemiology ontology, which will help further on the integration and communication among the community of epidemiologists.

2.2 Progress in the Reporting Period

- *A summary of progress towards objectives, and details for each task the team was involved;*
- *Highlight clearly significant results;*
- *Team publications within the scope of the project (please provide a pdf of the publication if possible);*
- *List of outreach activities (conferences, Invited talks, presentations, workshops, tutorials,);*
- *List of press releases or media coverage, any particular dissemination activity;*
- *unanticipated finding, opportunity etc*

Main Activities at FFCUL in the first semester of the fourth year of the project in WP3:

1. Mário Silva, João Zamite and Carlos Sousa attended the Epiwork Review in Brussels in March, 2012 – Mário Silva presented Deliverables D3.4 and D3.5 to the EC and gave an overview of the progress and challenges in WP3. A demo of the Epidemic Marketplace highlighting the improvements since the previous review was presented to the reviewers.
2. **[opportunity]** The LASIGE team established contacts with other teams involved in projects for the development of biomedical ontologies and terminologies. Miquel Porta, author of the popular “Dictionary of Epidemiology” is very interested in our development of an epidemiological semantic network to describe epidemiological datasets and a possible collaboration may happen. Miquel Porta, the editor of the Dictionary of Epidemiology, is contacting OUP for copyright issues.
3. Daniela Paolotti ISI was in Lisbon for a week in May. The discussions focused on the review of the pending issues with WP4-WP3 and WP5-WP3 integration. For the WP4-WP3 integration a release was planned. Regarding the WP5-WP3 integration, a version was already implemented in May, but resources are not being made public at the moment; the

workflow for the upload of public resources should be reviewed to create some "flagging" system of an upload that will create the necessary public policies. Discussed the possibility of having a meeting in Turin in September to finalize the integration of the WP3, WP4 and WP5 platforms.

4. MEDCollector datasets are all uploaded to the EM but need to be made public as well. The group also discussed with Daniela on the possibility and tactics for attracting more users to publish data on the EM. Also, the group discussed the possibility of having some feedback from Marco Quaggiotto and Wouter Van den Broeck on the EM interface to start improving its appeal.
5. **[opportunity]** Francisco Couto and João Ferreira met Lazlo Balkanyi (knowledge manager at ECDC) after our presentation of the NERO ontologies for epidemiology at the ICBO 2012. ECDC is developing a Terminology Services platform along the same lines as our NERO for Epiwork. We have made contact and intend to strengthen cooperation and develop synergies between both initiatives.
6. **[outreach]** The LASIGE team published and presented a full and a demo paper at the 2012 International Conference of Biomedical Ontologies (ICBO), which was a significant step towards the disclosure of our work to this community.

Publications and Presentations:

1. **[publication + presentation]** Francisco M. Couto, João D. Ferreira, João Zamite, Carlos Santos, Tiago Posse, Paulo Graça, Dulce Domingos, Mário J. Silva, The Epidemic Marketplace Platform: towards semantic characterization of epidemiological resources using biomedical ontologies. Demonstration at the International Conference on Biomedical Ontologies (ICBO) 2012.
2. **[publication + presentation]** João D. Ferreira, Catia Pesquita, Francisco M. Couto, Mário J. Silva, Bringing epidemiology into the Semantic Web. International Conference on Biomedical Ontologies (ICBO) 2012.

3. **[publication]** João D. Ferreira, Daniela Paolotti, Francisco M Couto, Mário J Silva. On the Usefulness of Ontologies in Epidemiology Research and Practice. Journal of Epidemiology and Community Health (in press).
4. **[publication]** Catia Pesquita, Francisco M. Couto 2012: Predicting the Extension of Biomedical Ontologies. PLOS Computational Biology 9(8), e1002630.
5. **[presentation]** João D. Ferreira & Francisco M. Couto. Title: Semantic Similarity in the Biomedical Domain. Poster presentation at the Bioinformatics Open Days 2012. Braga, Portugal.

Activities at FFCUL in the first semester of the fourth year of the project in Task 3.1:

- Completed the workflows required to fully support integration with WP4 (Upload, binary upload, delete resource, delete datastream, fetch datastream, search, list datastreams). The software will require optimisations and reliability should be increased in the next period.
- Established the protocols and procedures for integrating data harvested in WP5. LASIGE could not attend the WP5 meeting in Stockholm, where integration would be jointly tested. This has been partially successful with Skype calls between both groups, but will have to be completed in the next period.

Activities at FFCUL in the first semester of the third year of the project in Task 3.2:

- The team has researched strategies for the creation of a network of ontologies of epidemics (NERO), which could either involve the creation of a new ontology (as none of the existing initial proposals appears to have significant traction) or an automated conversion of the Dictionary of Epidemiology into a semantic network.
- Research on a second version of the metadata model. The existing is missing the mapping of ontologies of the terms currently used to annotate EM resources. Meanwhile, NERO is now integrated with the front-

end. The meta-data input boxes have an auto-complete facility showing the NERO terms that can be used for each field. The used ontologies will also be available as resources in the EM.

Activities at FFCUL in the first semester of the third year of the project in Task 3.3:

- A much improved front-end for access control has been developed, inspired on the model available on social network sites. The team worked on new access control schemes on social networks that could be adapted to the EM. These include the creation of dynamic user groups based on attributes (institutions or Facebook groups, or even friends/friends of friend's rules) that could be used for access control in the EM.
- Initiated the development of an improved curator interface. Discussions on the role of curators in the platform. This prompted the development of support to implement collections of resources as user-managed objects.

Activities at FFCUL in first semester of the third year of the project in Task 3.4:

- Continued work on the monitoring and log data collection and analysis tasks.

2.3 Effort Allocation

The effort allocated by LASIGE to WP3 in the reporting period (and previous years) is as follows:

WP3 Reporting Period	FFCUL	Effort (p.m)			
		Non-Perm	+	Perm	=Total
Year 1: February 1, 2009 to January 31, 2010		19.38	+	4.53	= 23.91
Year 2: February 1, 2010 to January 31, 2011		32.55	+	6.43	= 38.98
Year 3: February 1, 2011 to January 31, 2012		40.86	+	8.39	= 49.26
Year 4: February 1, 2012 to July 31, 2012		28.40	+	1.70	= 30.10

The commitment to the project in the first year was 23.91 persons.month, 19.38 p.m from technicians, and 4.53 p.m from permanent staff.

The commitment to the project in the second year was 38.98 persons.month, 32.55 p.m from technicians, and 6.43 p.m from permanent staff.

The commitment to the project in the third year was 49.26 persons.month, 40.86 p.m from technicians, and 8.39 p.m from permanent staff.

In the reporting period (1st semester of the 4th year), the commitment was 30.10 persons.month, 28.40 pm from technicians, and 1.70 pm from permanent staff.

The effort allocated is now well above planned (assuming a 10.25 p.m constant effort/semester), and reflects the decision of hiring more (and less skilled) staff for the FFCUL/LASIGE team of Epiwork. On the other hand, it also reflects an additional commitment by FFCUL/LASIGE to the project to address the weaknesses resulting from having hired less skilled researchers.

2. Work in WP4 — Epidemic Modelling Platform

This task is lead by ISI, with a total FFCUL contribution of 19 persons.month (12 hired + (7) academics).

Activities at FFCUL in the seventh semester of the project in WP4:

The LASIGE work on WP4 is related to the integration of the Computational and Data Platforms of Epiwork, which has been accounted in detail in the description of activities of LASIGE on WP3. Below, we mention the most significant of these activities:

- Mário Silva, João Zamite and Carlos Sousa attended the Epiwork Review in Brussels in March, 2012. A joint demo of the integration of the Epidemic Marketplace and Gleanviz was made at the review.
- Daniela Paolotti ISI was in Lisbon for a week in May. The discussions focused on the review of the pending issues with WP4-WP3 and WP5-WP3 integration. Discussed the possibility of having a meeting in Turin in September to finalize the integration of the WP3, WP4 and WP5 platforms.
- ISI and LASIGE members met regularly by videoconference on a weekly basis during the period.

2.4 Effort Allocation

The effort allocated to WP4 in the in the reporting period (and previous years) is as follows:

WP4 Reporting Period	FFCUL	Effort (p.m)		
		Non-perm	+	Perm
		=Total		

WP4 Reporting Period	FFCUL	Effort (p.m) Non-perm + Perm =Total
Year 1: February 1, 2009 to January 31, 2010		5.72+0.50 = 6.22
Year 2: February 1, 2010 to January 31, 2011		5.05 + 0.40 = 5.45
Year 3: February 1, 2011 to January 31, 2012		7.31 + 2.53 = 9.84
Year 4: February 1, 2012 to July 31, 2012		2.79 + 2.37 = 5.16

The commitment to the project in the first year was 6.22 persons.month, 5.72 p.m from technicians, and 0.50 p.m from permanent staff.

The commitment to the project in the second year was 5.45 persons.month, 5.05 pm from technicians, and 0.40 p.m from permanent staff.

The commitment to the project in the third year was 9.84 persons.month, 7.31 pm from technicians, and 2.53 p.m from permanent staff.

In the reporting period (1st semester of the 4th year), the commitment was 5.16 persons.month, 2.79 pm from technicians, and 2.37 from permanent staff.

This is as planned, reflecting the fact that this task is led by ISI and our contribution is intensifying in the second half of the project, as we settle on a common software architecture, including interfaces, for orchestrating and running services on both platforms. The interactions have now been strengthened to weekly teleconference meetings, as we proceed on the provision of an API suited to the requirements of the computational platform.

Work in WP7 — Management

This task is lead by ISI.

FFCUL Effort in this task: 4 persons.month.

Activities at FFCUL in the fifth semester of the project in WP7:

1. Data collection activities for project tracking, preparation of the 7th semester report, management presentations.
2. Mário Silva attended the March 2012 Project Review in Brussels.

2.5 Effort Allocation

The effort allocated to WP7 in the in the reporting period (and previous years) is as follows:

WP7 Reporting Period	FFCUL	Effort (p.m) Non-Perm + Perm = Total
Year 1: February 1, 2009 to January 31, 2010		0.0 + 1.10 = 1.10
Year 2: February 1, 2010 to January 31, 2011		0.0 + 0.96 = 0.96
Year 3: February 1, 2011 to January 31, 2012		0.0 + 0.99 = 0.99
Year 3: February 1, 2012 to July 31, 2012		0.0 + 0.23 = 0.23

The effort dedicated to the project in the period, 0.23 p.m., was 100% contributed by permanent staff, as planned.